Introduction to FIFE
Getting the Experiments Computing in Tune

Mike Kirby

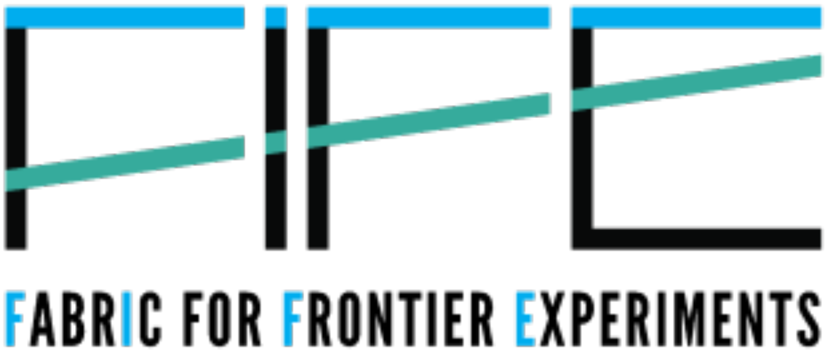
Linux at Fermilab Meeting

July 27, 2016

FABRIC FOR FRONTIER EXPERIMENTS

FIFE Logo

- Just design it at home (or maybe my bosses kid)



- Maybe send the wrong message

FIFE
Computing for Experiments

- Maybe copyright problems (but not really)

Introduction to FIFE

- The **F**abr**I**c for **F**rontier **E**xperiments aims to
- Lead the development of the computing model for non-LHC experiments
- Provide a robust, common, modular set of tools for experiments, including
  - Job submission, monitoring, and management software
  - Data management and transfer tools
  - Database and conditions monitoring
  - Collaboration tools such as electronic logbooks, shift schedulers
- Work closely with experiment contacts during all phases of development and testing
- https://web.fnal.gov/project/FIFE/SitePages/Home.aspx

A Wide Variety of Stakeholders

- At least one experiment in energy, intensity, and cosmic frontiers, studying all physics drivers from the P5 report, uses some or all of the FIFE tools (massive neutrino presence)
- A wide variety of computing models (1980s-era to future experiments); FIFE tools are adaptable to them all

CDF II

Main Injector Neutrino Oscillation Search

Mu2e

DARKSIDE

SeaQuest
E906

g-2

µBooNE

LArIAT

SHORT-BASELINE SBND NEAR DETECTOR

NOvA

THE DARK ENERGY SURVEY

DUNE DEEP UNDERGROUND NEUTRINO EXPERIMENT

Common problems, common solutions

- FIFE experiments on average are 1-2 orders of magnitude smaller than LHC experiments; often lack sufficient expertise or time to tackle all problems, e.g. software frameworks or job submission tools
  - Very common to be on multiple experiments in the neutrino world - **familiarity with FIFE has been extremely successful as people move from one to another**
- By bringing experiments under a common umbrella, can leverage each other's expertise and lessons learned
- Common software frameworks are also available (ART, based on CMSSW) for most experiments
- FIFE also provides a voice within the larger community
  - –active part of the OSG and HEPCloud; contribute to toolset
  - –provide access to computing resources not readily available to all experiments (OSG, Condor, ASCR, NERSC, etc)

Centralized Services from FIFE

•Submission to distributed computing – JobSub, GlideinWMS frontend

•Processing Monitors, Alarms, and Automated Submission

•Data Handling and Distribution

–Sequential Access Via Metadata (SAM) File Transfer Service

–dCache/Enstore

–Intensity Frontier Data Handling Client

•Software stack distribution – CERN Virtual Machine File System (CVMFS)

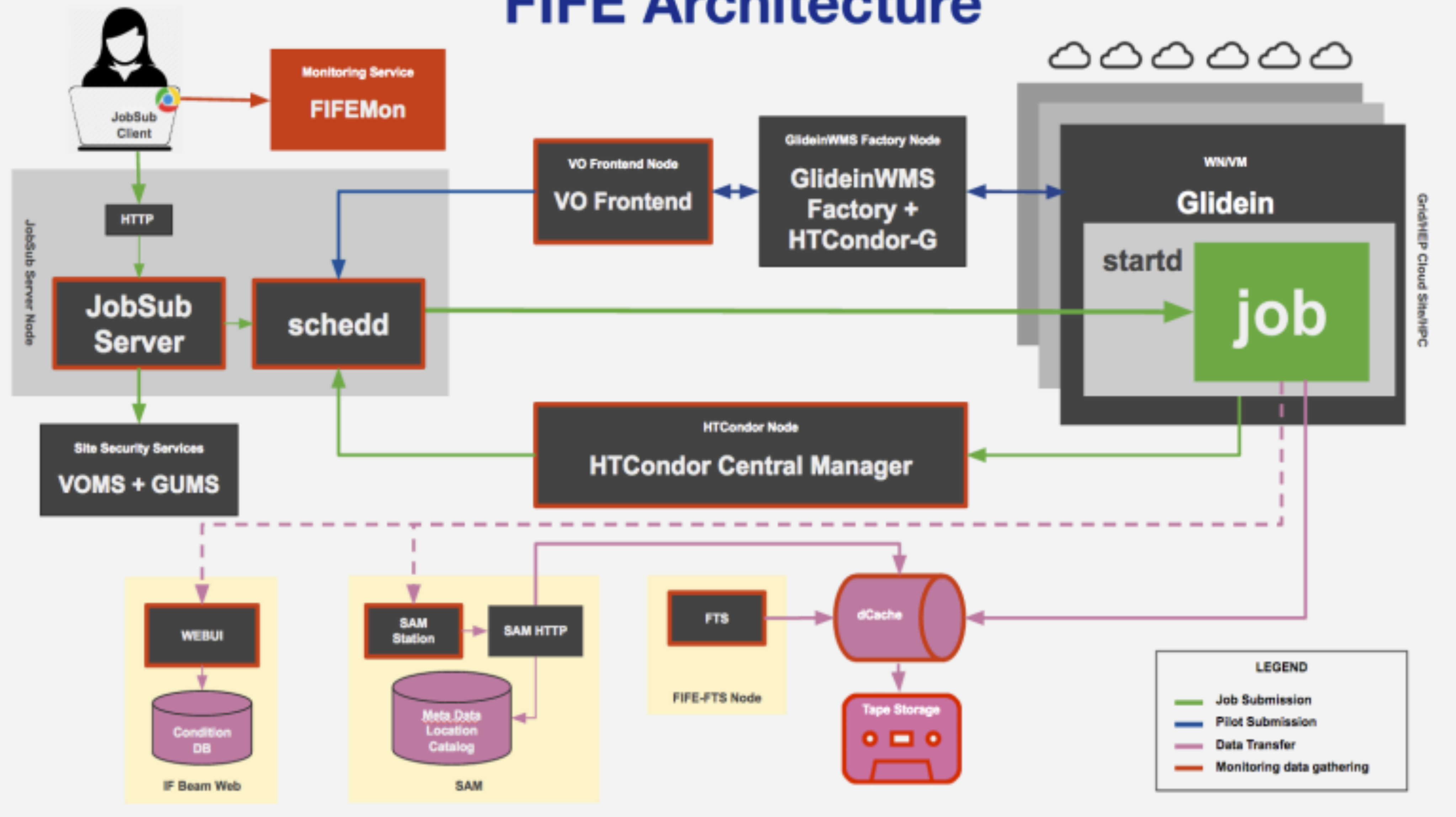
•User Authentication, Proxy generation, and security

•Electronic Logbooks, Databases, and Beam information

•Integration with future projects, e.g. HEPCloud

FABRIC FOR FRONTIER EXPERIMENTS

FIFE Architecture



JobSub Client

Monitoring Service
FIFEMon

JobSub Server Node

HTTP

JobSub Server

schedd

Site Security Services
VOMS + GUMS

VO Frontend Node
VO Frontend

GlideinWMS Factory Node
GlideinWMS Factory + HTCondor-G

WN/VM
Glidein

startd

job

Grid/HEP Cloud Site/HPC

HTCondor Node
HTCondor Central Manager

WEBUI

Condition DB

IF Beam Web

SAM Station

SAM HTTP

Meta Data Location Catalog

SAM

FTS

FIFE-FTS Node

dCache

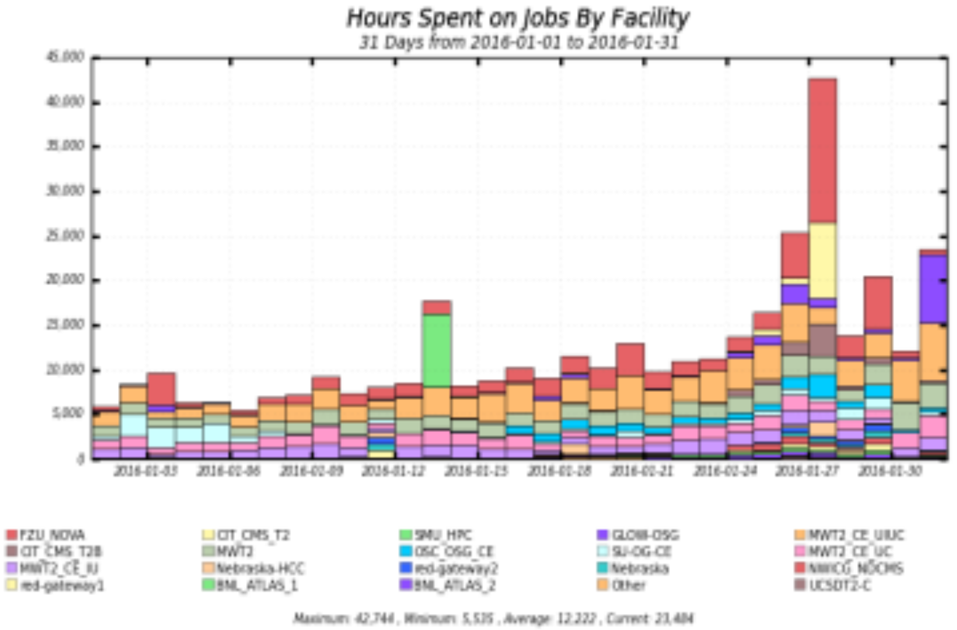
Tape Storage

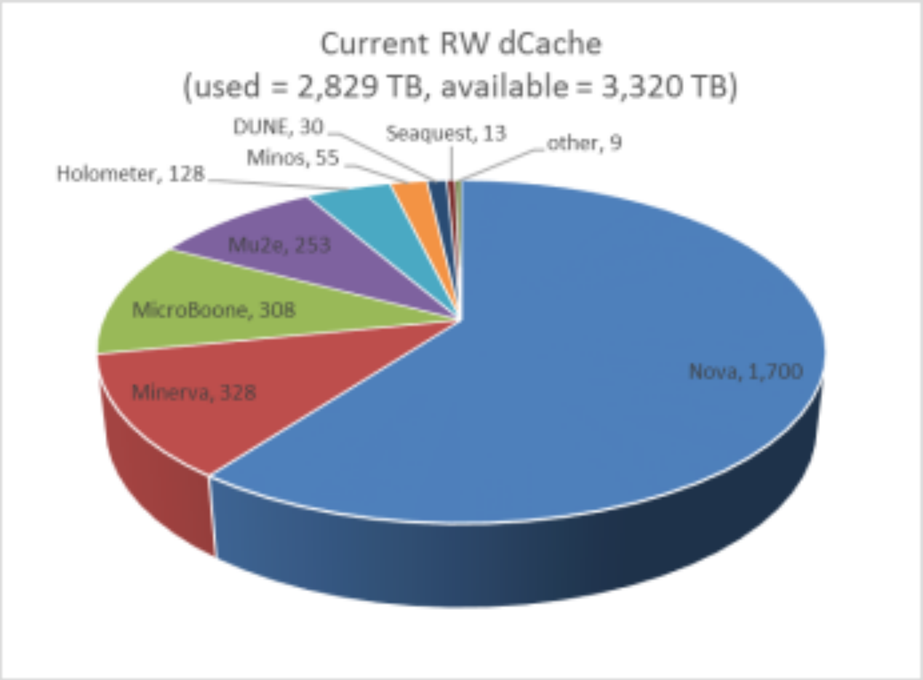
LEGEND

🟩	Job Submission
🟦	Pilot Submission
🟪	Data Transfer
🟥	Monitoring data gathering

NOvA – full integration of FIFE Services

- File Transfer Service stored 1.7 PB of NOvA data in dCache and Enstore
- SAM Catalog contains more than 41 million files
- Helped develop SAM4Users as lightweight catalog

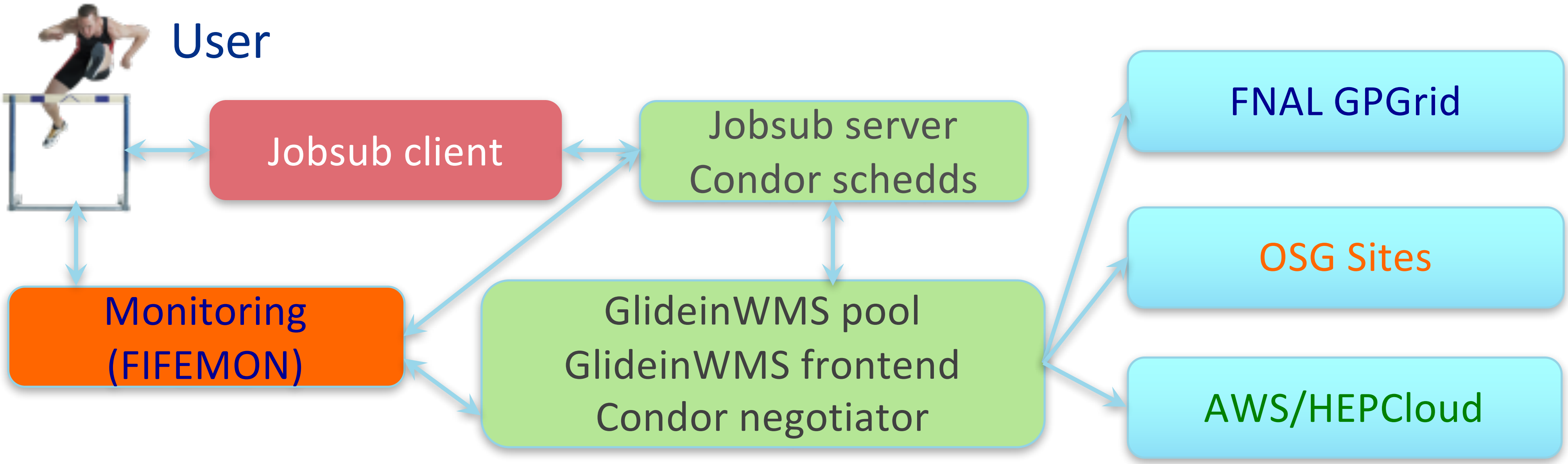




- Jan 2016 - NOvA published first papers on oscillation measurements
- avg 12K CPU hours/day on remote resources
- > 500 CPU cores opportunistic
- FIFE group enabled access to remote resources and helped configure software stack to operate on remote sites
- Identified inefficient workflows and helped analyzers optimize

Job Submission and management architecture

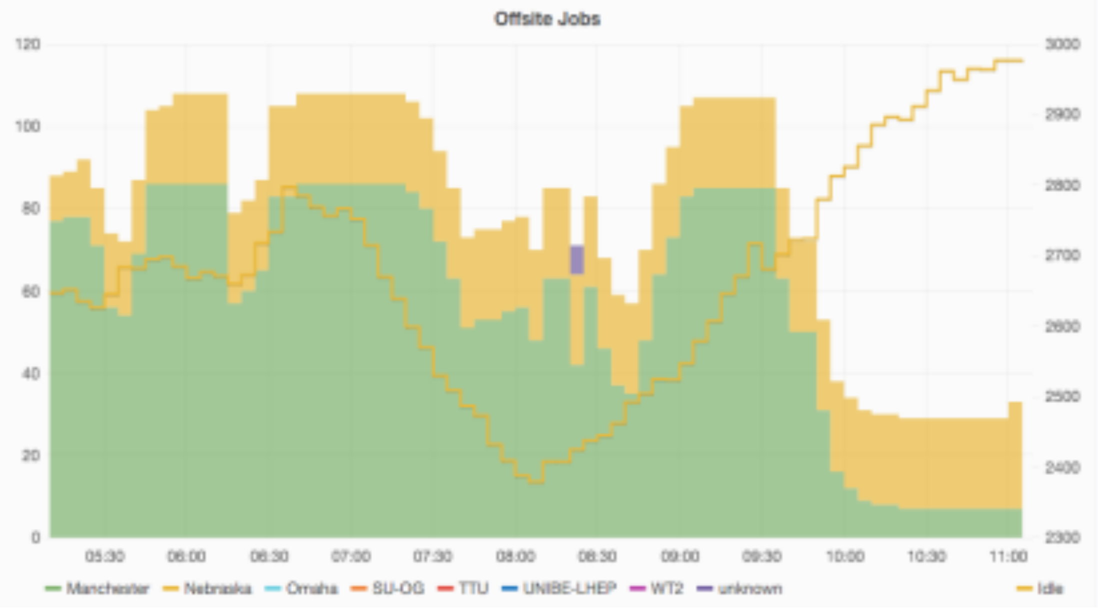
- Common infrastructure is the **fifebatch** system: one GlideInWMS pool, 2 schedds, frontend, collectors, etc.
- Users interface with system via “jobsub”: middleware that provides a *common tool across all experiments*; shields user from intricacies of Condor
  - Simple matter of a command-line option to steer jobs to different sites
- Common monitoring provided by FIFEMON tools
  - Now also helps users to understand why jobs aren’t running



New International Sites for running jobs

- Previously had allocation for NOvA at FZU in Prague
- Have since added Manchester, Lancaster, and Bern for Microboone (only) in recent weeks
  - Alessandra Forti very helpful at Manchester; Gianfranco Sciacca at Bern; Matt Doidge at Lancaster
- Setup in both cases was about one week in both cases
  - Lancaster integration was < 1 week

Made so smooth by GWMS and OSG's ongoing work to make variety of different sites compatible; **same can easily be done for DUNE as new sites come online**



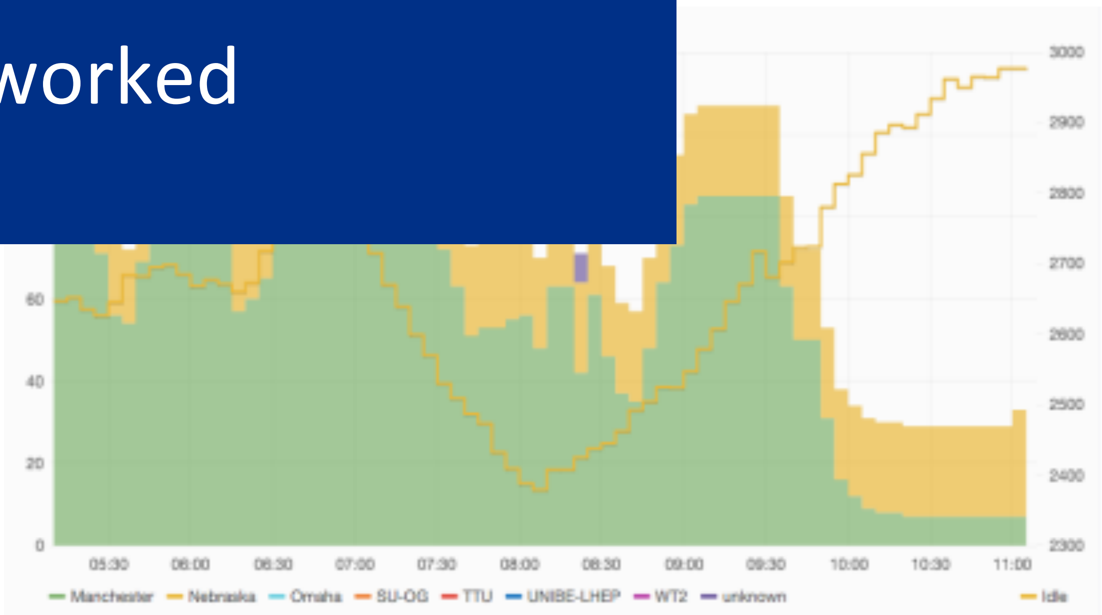
New International Sites for running jobs

- Previously had allocation for NOvA at FZU in Prague
- Have since added Manchester, Lancaster, and Bern for Microboo
  - Alessan nfranco Sciacca
- Setup in b th cases
  - Lancast

Made so smoot GWMS and OS work to make v different sites compatible; **same can easily be done for DUNE as new sites come online**

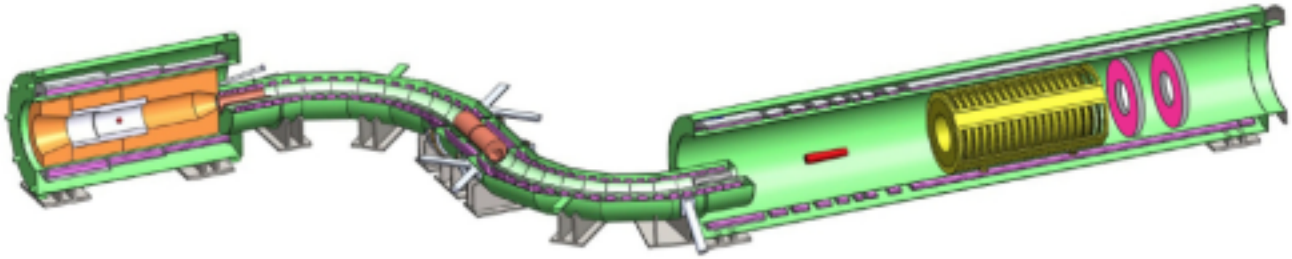
How “smooth” was it really?

The very first test jobs at Manchester and Lancaster worked



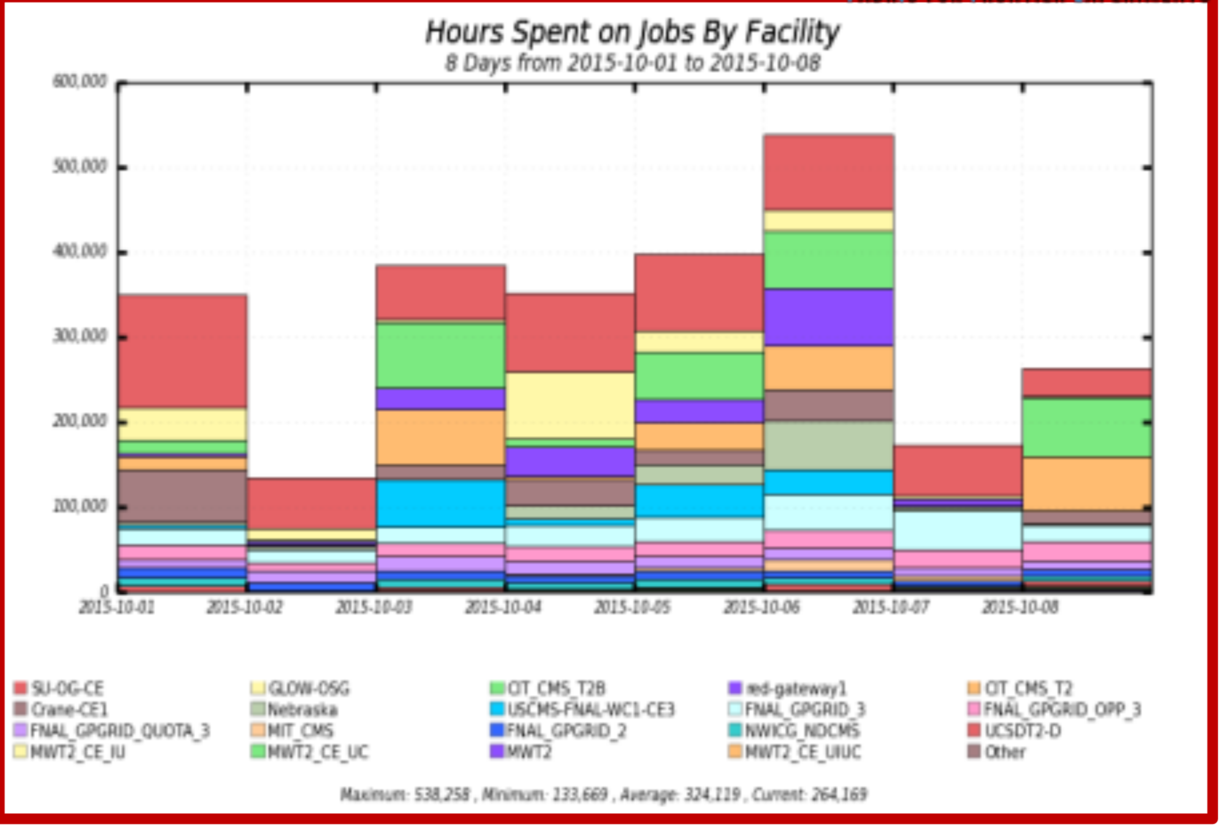
Mu2e Beam Simulations Campaign

- Mu2e recently received CD3 approval – review design of beam transport, magnets, detectors, and radiation
- Approval required a combination of beam intensity and magnet complexity, necessitated significant simulation studies
  - –12 Million CPU hours in 6 months estimate for required precision
- Well beyond the available resources at Fermilab allocated to Mu2e
- FIFE support group helped deploy Mu2e beam simulation software stack through CVMFS to remote sites
- Helped probe additional remote resources and integrate into job submission – ideally without user knowledge



Mu2e Beam Simulations Campaign

- Almost no input files
- Heavy CPU usage
- < 100 MB output
- Ran > 20M CPU-hours in under 5 months
- Avg 8000 simultaneous jobs across > 15 remote sites



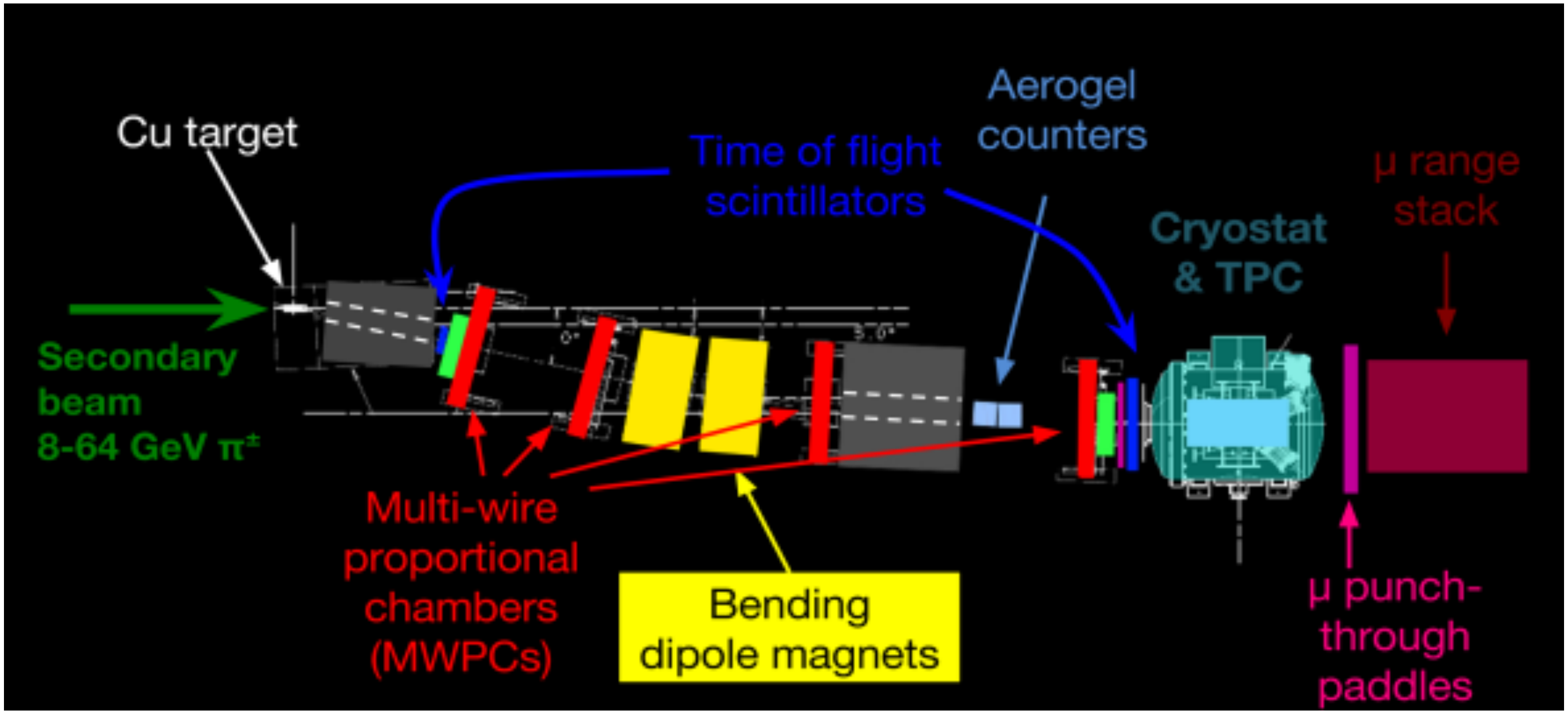
- Usage as high as 20,000 simultaneous jobs and 500,000 CPU hours in one day – peaked usage 1st wk Oct 2015
- *Achieved stretch goal* for processing 24 times live-time data for 3 most important backgrounds
- **Total cost to Mu2e for these resources: $0**

Recent challenges for FIFE Experiments

- Code distribution via CVMFS generally works very well
  - Differences in installed software on worker nodes causes occasional problems (mostly X11 libs, i.e. things users assume are always installed)
  - Helped experiments work around this by creating packages of libraries within CVMFS
- Memory requirements
  - Younger experiments (particularly LAr TPC expts.) have workflows requiring > 2 GB memory per job. Somewhat limited resources available going above 2 GB/1 core.
- Large auxiliary files
  - StashCache looking promising; helping develop and test the tools
- Data management for users

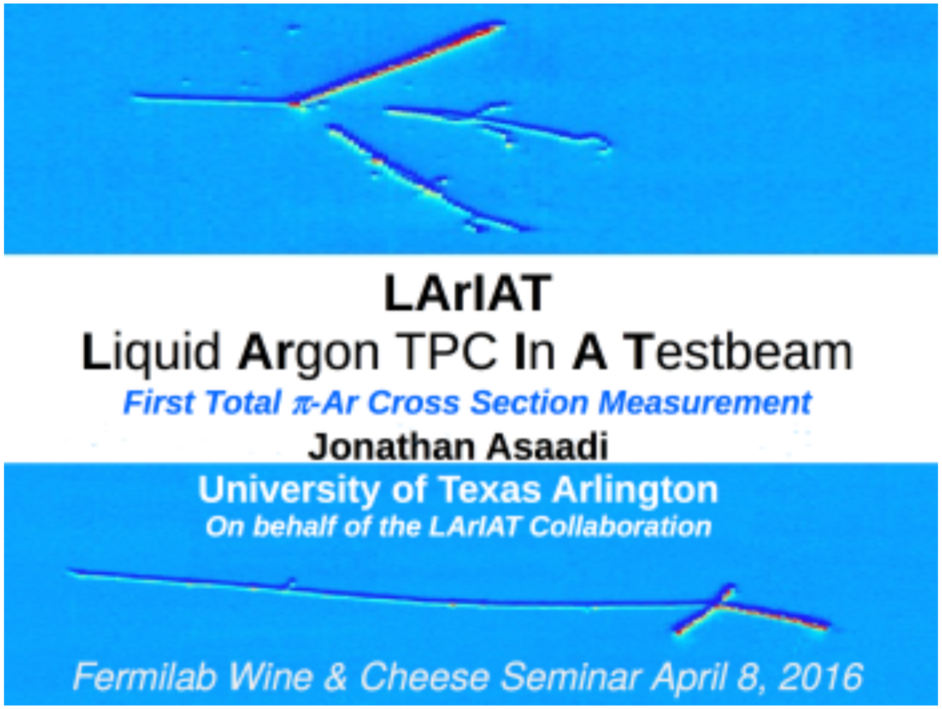
Enhancement of LArIAT SAM File catalog

- Liquid Argon In A Testbeam - exploring the cross-sections on LAr for final state particles
- Important for understanding the response in future detectors
- Incident beam can change every day, but DAQ not coupled to bending magnets – incorporate beam db into file catalog



Graphic courtesy
Of Dr. Jason St. John

Enhancement of LArIAT SAM File Catalog



LArIAT

Liquid Argon TPC In A Testbeam

First Total π-Ar Cross Section Measurement

Jonathan Asaadi

University of Texas Arlington

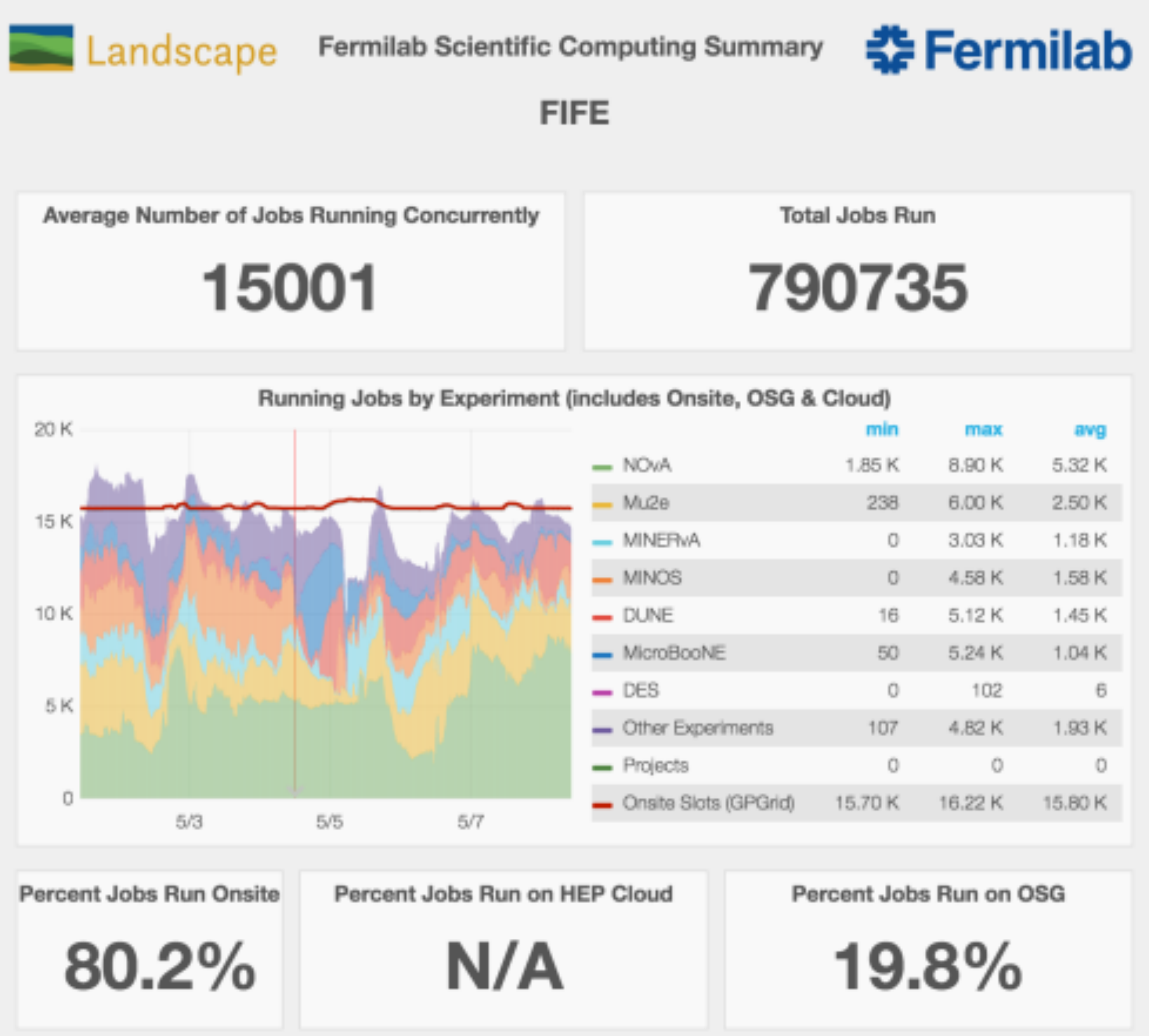
On behalf of the LArIAT Collaboration

Fermilab Wine & Cheese Seminar April 8, 2016

- Extended the capability of SAM to be able to interface with external databases
- Allows for LArIAT to select data based upon criteria from the beam condition database
- DAQ and Offline processing are independent of beam database so that this is not a blocking situation
- FIFE Support team helped to instantiate and configure this beam db integration with LArIAT SAM Catalog
- Analyzers focused on physics instead of computing
- LArIAT presented first cross-sections at W&C April 8, 2016

FIFE Monitoring of resource utilization

- Extremely important to understand performance of system
- Critical for responding to downtimes and identifying inefficiencies
- Focused on improving the real time monitoring of distributed jobs, services, and user experience



Landscape Fermilab Scientific Computing Summary Fermilab

FIFE

Average Number of Jobs Running Concurrently

15001

Total Jobs Run

790735

Running Jobs by Experiment (includes Onsite, OSG & Cloud)

	min	max	avg
NOvA	1.85 K	8.90 K	5.32 K
Mu2e	238	6.00 K	2.50 K
MINERvA	0	3.03 K	1.18 K
MINOS	0	4.58 K	1.58 K
DUNE	16	5.12 K	1.45 K
MicroBooNE	50	5.24 K	1.04 K
DES	0	102	6
Other Experiments	107	4.82 K	1.93 K
Projects	0	0	0
Onsite Slots (GPGrid)	15.70 K	16.22 K	15.80 K

Percent Jobs Run Onsite

80.2%

Percent Jobs Run on HEP Cloud

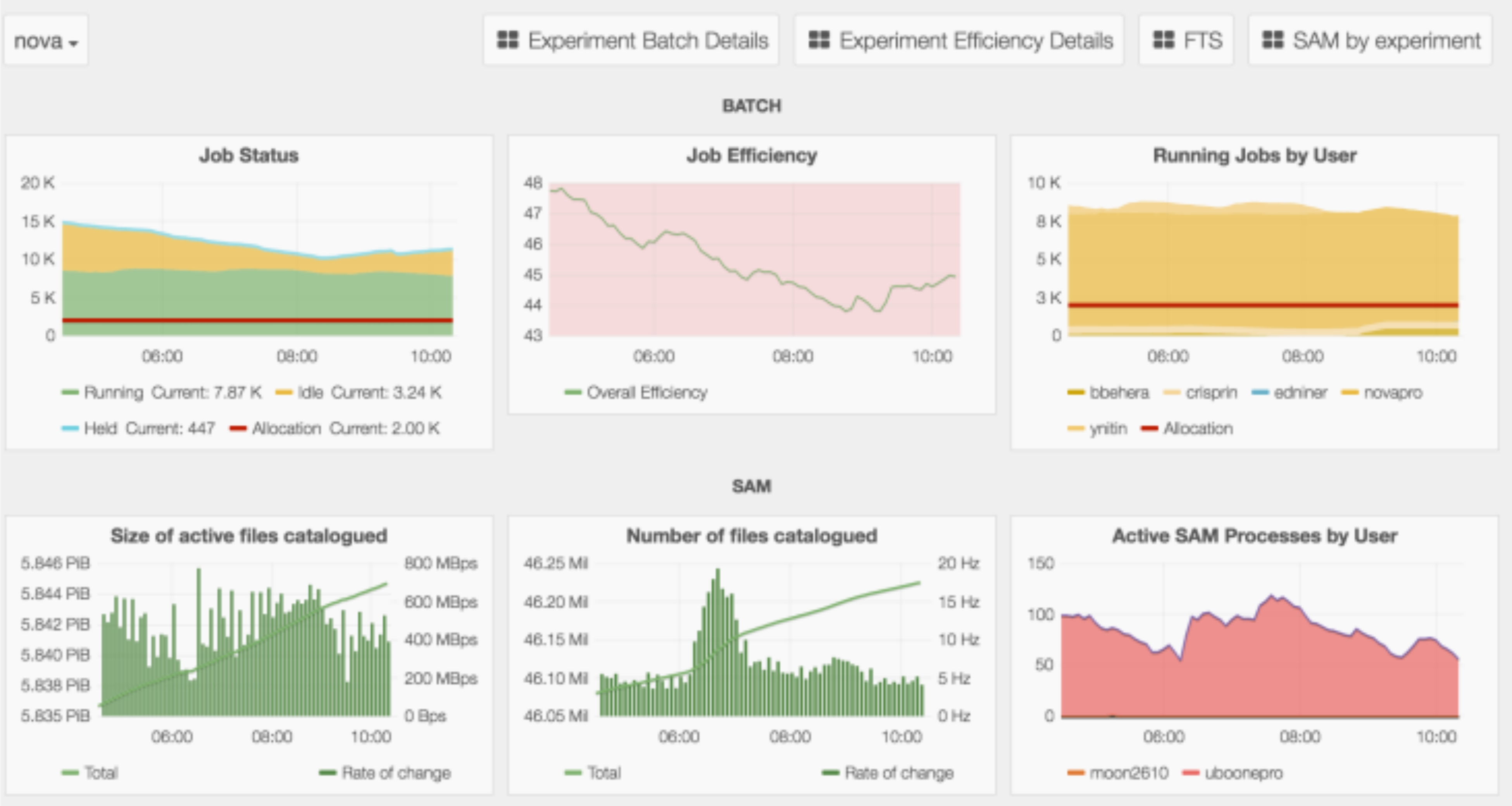
N/A

Percent Jobs Run on OSG

19.8%

Detailed profiling of experiment operations

FABRIC FOR FRONTIER EXPERIMENTS



Allows identification for inefficiencies, potential slow downs, or blocking conditions in workflows

Fermilab

Production Management: POMS

Developing system to full manage entire production workflow: POMS

POMS can currently:

Track what processing needs to be done (“Campaigns”)

Track job submissions made for above

Automatically make job submissions for above

Launch recovery jobs for files that didn't process automatically

Launch jobs for dependent campaigns automatically to process output of previous passes.

Interface with SAM to track files processed by submissions and Campaigns

Provides “Triage” interface for examining individual jobs/logs and debugging failures.

POMS Configuration

Telling POMS about your software and scripts is done through a 5tiered configuration system

Experiment name and users added to POMS (by admins)

Launch Template login and setup info to run jobs (also adding POMS special principal to appropriate .k5login files)

“Campaign Definition” for types of jobs you run how to launch a MonteCarlo job, or a Reconstruction job,etc.

Specific Campaign we want to run Reconstruction on these three specific datasets...

You can also configure what types of recovery jobs should be run, and what campaigns depend on others.

Full details in Anna’s talk

FIFE Plans for the future

- Increase use of POMS among experiments across all frontiers
  - Goal is to automate production as much as possible, eliminate need for experiments to create the infrastructure
- Help define the overall computing model of the future by seamlessly integrating dedicated, opportunistic, and commercial computing resources via HEPCloud
  - Increase access to HPC resources for job submission
  - Usher in easy access to GPU resources for those experiments interested (Minerva, NOvA, uboone, DUNE, etc.)
  - Lower barriers to accessing computing elements around the world in multiple architectures
- Scale up and improve UI to existing services

Backup

FSAM and FFTS

Fermi SAM is an interweaving of several things

- A File metadata/provenance catalog
- A File replica catalog
- Allows metadata query based “dataset” creation
- An optimized “project” File delivery system

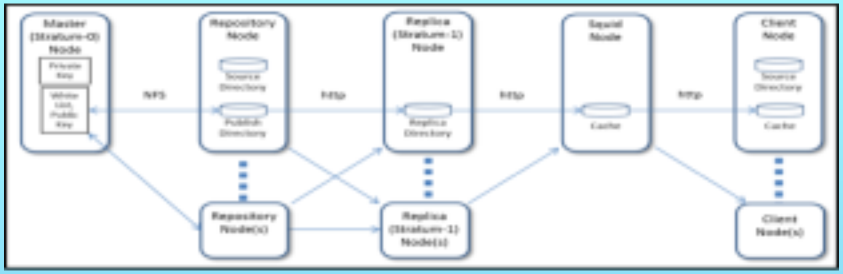
Fermi File Transfer Service

- Watches one or more dropboxes for new files
- Can extract metadata from files and declare to FSAM, or handle files already declared
- Copies files to one or more destinations based on file metadata and/or dropbox used
- registers/unregisters file locations in FSAM
- Cleans dropboxes, usually N days after files are on tape

Contributing back to software stack

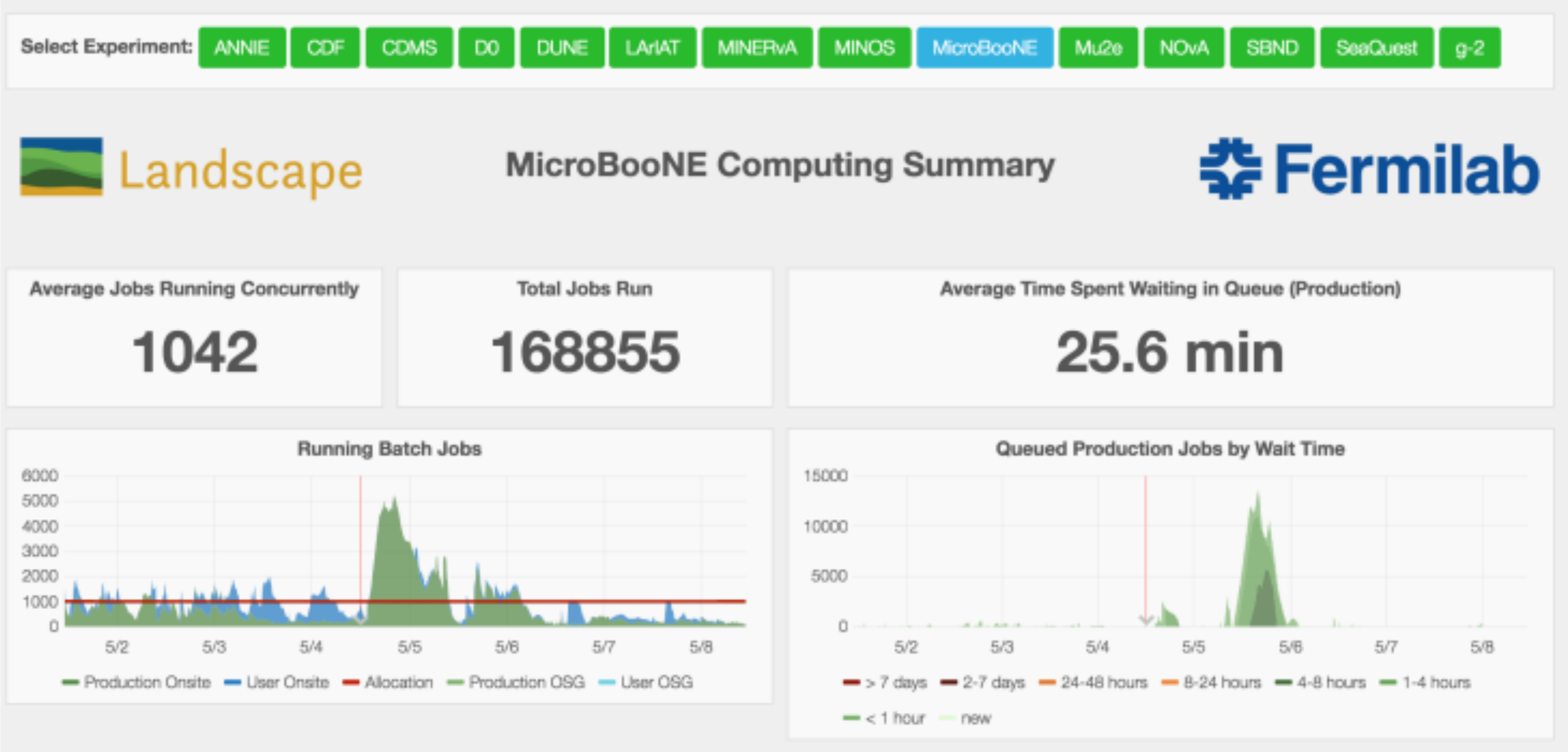
THE SOFTWARE DEVELOPMENT CYCLE
1 PLANNING
2 ANALYSIS
3 DESIGN
4 IMPLEMENTATION
5 TESTING & INTEGRATION
6 MAINTENANCE

Software distribution to Grid worker nodes: CVMFS infrastructure



- increase of Fermilab experiments utilizing OASIS CVMFS caused conflicts updating and syncing software on OASIS
- To relieve conflicts Fermilab worked with CERN to update CVMFS and OASIS to integrate remote CVMFS repositories
- CVMFS repositories located at sites (Fermilab, other labs)
- distribution of large files for simulation tasks -> development of StashCache
- FIFE served the role of collating and communicating requirements, and contributing to design, testing, and implementation to include monitoring and tracking usage

Overview of Experiment Computing Operations



MicroBooNE Computing Summary

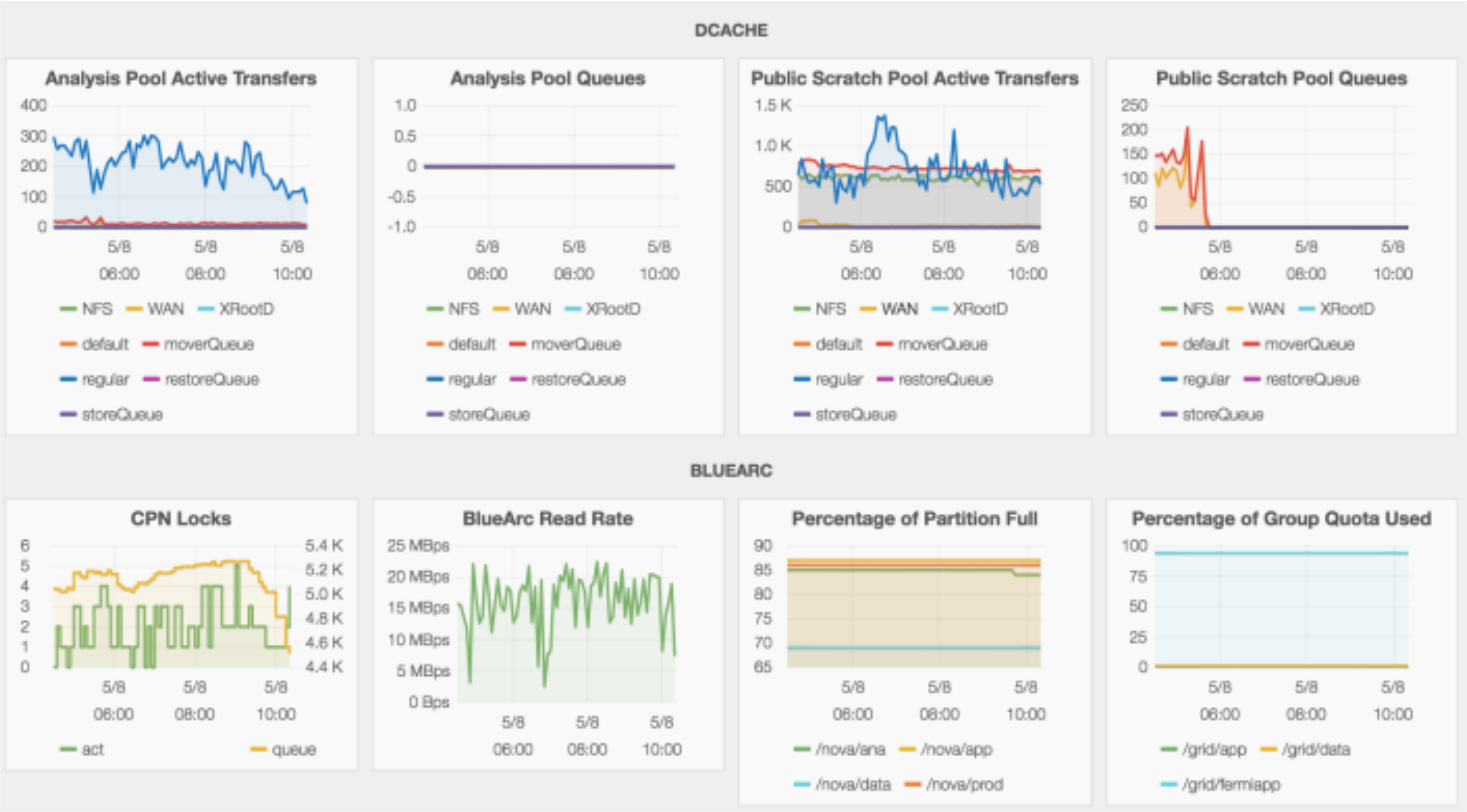
Average Jobs Running Concurrently: 1042

Total Jobs Run: 168855

Average Time Spent Waiting in Queue (Production): 25.6 min

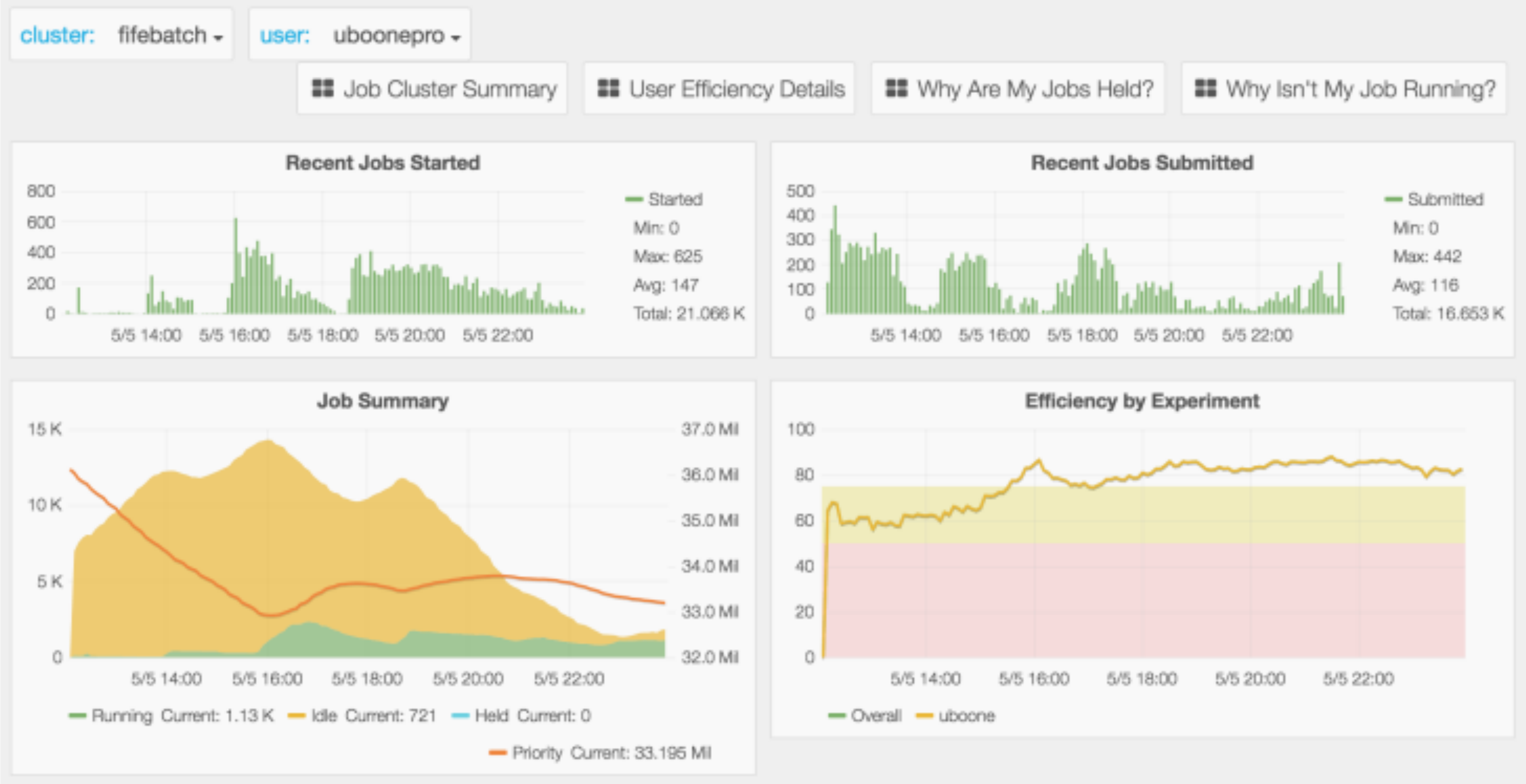
quickly understand the usage pattern for the last week of each experiment and collectively get a picture of distributed computing operations for the FIFE experiments

Detailed profiling of experiment operations



Monitor usage of slow moving resources so that projections can be made for projecting future need and limitations

Monitoring of jobs and experimental dashboards

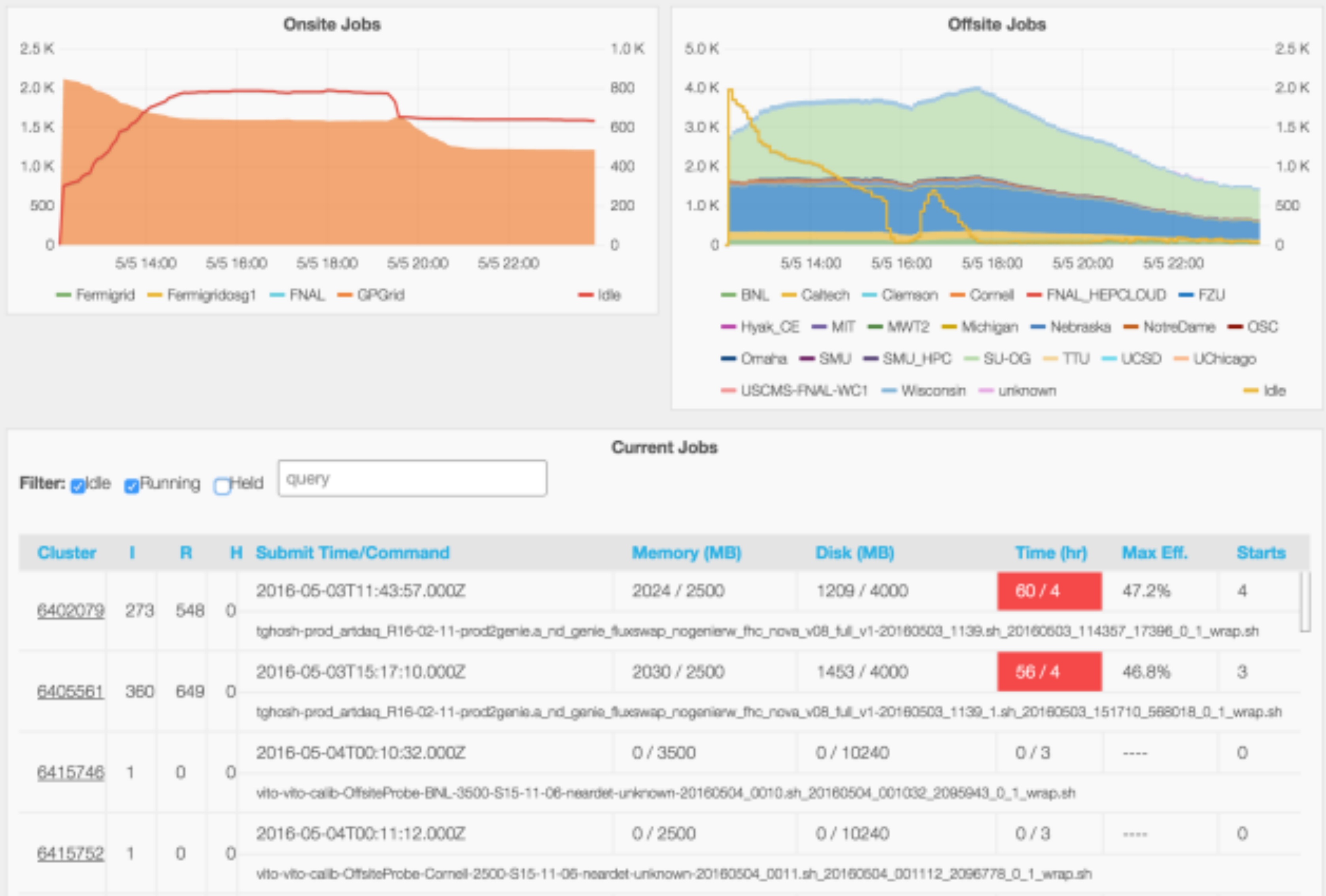


cluster: fifebatch

user: uboonepro

Monitoring for individual users to track their distributed computing workflows and understand their resource allocation and needs

Monitoring of jobs and experiment dashboards



Current Jobs

Filter: ☑Idle ☑Running ☐Held query

Cluster	I	R	H	Submit Time/Command	Memory (MB)	Disk (MB)	Time (hr)	Max Eff.	Starts
6402079	273	548	0	2016-05-03T11:43:57.000Z tghosh-prod_artdaq_R16-02-11-prod2genie.a_nd_genie_fluxswap_nogenierw_fhc_nova_v08_full_v1-20160503_1139.sh_20160503_114357_17396_0_1_wrap.sh	2024 / 2500	1209 / 4000	60 / 4	47.2%	4
6405561	360	649	0	2016-05-03T15:17:10.000Z tghosh-prod_artdaq_R16-02-11-prod2genie.a_nd_genie_fluxswap_nogenierw_fhc_nova_v08_full_v1-20160503_1139_1.sh_20160503_151710_568018_0_1_wrap.sh	2030 / 2500	1453 / 4000	56 / 4	46.8%	3
6415746	1	0	0	2016-05-04T00:10:32.000Z vito-vito-calib-OffsiteProbe-BNL-3500-S15-11-06-neardet-unknown-20160504_0010.sh_20160504_001032_2095943_0_1_wrap.sh	0 / 3500	0 / 10240	0 / 3	----	0
6415752	1	0	0	2016-05-04T00:11:12.000Z vito-vito-calib-OffsiteProbe-Cornell-2500-S15-11-06-neardet-unknown-20160504_0011.sh_20160504_001112_2096778_0_1_wrap.sh	0 / 2500	0 / 10240	0 / 3	----	0